Expected Genetic Contributions and Their Impact on Gene Flow and Genetic Gain

Expected Genetic Contributions and Their Impact on Gene Flow and Genetic Gain

J. A. Woolliams,* P. Bijma† and B. Villanueva‡

**Roslin Institute (Edinburgh), Roslin, Midlothian EH25 9PS, United Kingdom, †Animal Breeding and Genetics Group, Wageningen Institute of Animal Sciences, Wageningen Agricultural University, 6700 A4 Wageningen, The Netherlands and ‡Scottish Agricultural College, Edinburgh EH9 3JG, United Kingdom*

Manuscript received November 10, 1998
Accepted for publication June 14, 1999

ABSTRACT

Long-term genetic contributions (r_i) measure lasting gene flow from an individual *i*. By accounting for linkage disequilibrium generated by selection both within and between breeding groups (categories), assuming the infinitesimal model, a general formula was derived for the expected contribution of ancestor *i* in category *q* ($\mu_{i(q)}$), given its selective advantages ($\boldsymbol{s}_{i(q)}$). Results were applied to overlapping generations and to a variety of modes of inheritance and selection indices. Genetic gain was related to the covariance between r_i and the Mendelian sampling deviation (a_i), thereby linking gain to pedigree development. When $\boldsymbol{s}_{i(q)}$ includes a_i, gain was related to $E[\mu_{i(q)}a_i]$, decomposing it into components attributable to within and between families, within each category, for each element of $\boldsymbol{s}_{i(q)}$. The formula for $\mu_{i(q)}$ was consistent with previous index theory for predicting gain in discrete generations. For overlapping generations, accurate predictions of gene flow were obtained among and within categories in contrast to previous theory that gave qualitative errors among categories and no predictions within. The generation interval was defined as the period for which $\mu_{i(q)}$, summed over all ancestors born in that period, equaled 1. Predictive accuracy was supported by simulation results for gain and contributions with sib-indices, BLUP selection, and selection with imprinted variation.

SELECTION theory has not generally addressed how the number of descendants from an individual grows or reduces over time in relation to properties of the population. This is perhaps surprising because the development of the pedigree over generations provides the framework for the passage of genes through the population, forming the link between our understanding of individual genotypes and the way such genotypes influence the population. Such an understanding provides answers to, for example, the relative importance of individuals within a generation; where genetic change has arisen; how quickly the change generated has spread through the population; with what precision we are able to predict this change; how genetic change is related to the loss of variation; and how genetic change in one generation relates to that in a subsequent generation. These questions have no general framework within which they can be answered although some special cases have been investigated (*e.g.*, Villanueva *et al.* 1996; Bijma and Woolliams 1999).

The objective of this study is to describe the expectations for the proliferation of genetic lines using the concept of genetic contributions. The generation of linkage disequilibrium during selection changes the impact of selective advantages and this must be accounted for to predict the flow of an individual's genes through a population over time. These changes affect the comparative gene flow of different breeding groups or categories and of different individuals within categories. The general development builds upon the pioneering work of Wray and Thompson (1990) and more recently the studies of Woolliams *et al.* (1993; mass selection), Wray *et al.* (1994; sib-indices), and Woolliams and Thompson (1994). First, the concept of genetic contributions is considered in relation to genetic gain, and a general formula for gain is proved. The expected genetic contribution of an individual to subsequent generations is derived, and the relationship of the long-term genetic contribution with gain is used to show the consistency between the developed theory and classical theory (*e.g.*, Bulmer 1980). The concept of the generation interval is reevaluated as a natural extension of the contribution theory. Many of the detailed results are derived assuming an equilibrium. The uses of the developed formulae are shown in examples of selection applied to discrete generations using sib-indices, using best linear unbiased predictors (BLUP), with imprinted variation, and with overlapping generations.

Corresponding author: J. A. Woolliams, Roslin Institute (Edinburgh), Roslin, Midlothian EH25 9PS, United Kingdom.
E-mail: john.woolliams@bbsrc.ac.uk

MATERIALS AND METHODS

Definitions and basic notation: Table 1 shows the notation for the principal parameters. The concept of genetic contribu-

TABLE 1

The notational conventions for the principal parameters

t, u	Time variables
p, q	One of a total of n_c categories defined by sex and age
i, j, $i(q)$	Individuals in the population; $i(q)$ denotes individual in category q
G_t, ΔG	Genetic merit of population at time t and rate of genetic gain
T_m, T_f	Number of male and female candidates available for selection
$r_{i(q)}$	Long-term contribution of i in category q
a_i, A_i	Mendelian sampling term and breeding value of i
S_i	Selection score for i; 0 or 1 according to i selected or not
$\boldsymbol{s}_{i(q)}$	Vector of selective advantages for $i(q)$, of length n_s; mean over all selected in category q is $\bar{\boldsymbol{s}}_q$
$\mu_{i(q)}$	Expected long-term contribution, assumed to be linear regression on $\boldsymbol{s}_{i(q)}$ of form $\alpha_{i(q)} + \beta_q^T(\boldsymbol{s}_{i(q)} - \bar{\boldsymbol{s}}_q)$
α, β	Vectors of the coefficients for $\mu_{i(q)}$, of lengths n_c and $n_c\, n_s$, respectively
X_p	Number of parents in category p; the $n_c \times n_c$ diagonal matrix $\boldsymbol{N}$ has elements X_p
g_{pq}	Proportion of genes of selected individuals in category p that derive from parents of category q; the $n_c \times n_c$ matrix $\boldsymbol{G}$ has elements g_{pq}; the $n_c \times n_c$ matrix $\boldsymbol{G}_p$ has elements other than the pth row equal to zero.
$g_{0,\,pq}$	Proportion of genes among the newborn from which category p are selected that derive from parents of category q
λ_{pq}	Regression coefficients of proportion selected in category p on $\boldsymbol{s}_{i(q)}$ for parent in category q; has dimension $1 \times n_s$; the $n_c \times n_c n_s$ matrix Λ has elements λ_{pq}
π_{pq}	Regression coefficients of $\boldsymbol{s}_{j(p)}$ on $\boldsymbol{s}_{i(q)}$ for parent in category q; has dimension $n_s \times n_s$; the $n_c n_s \times n_c n_s$ matrix Π has elements π_{pq}
h^2, h_0^2	Heritability of trait in candidates, and heritability in unselected base generation
σ_A^2, σ_P^2, σ_I^2	Additive genetic, phenotypic, and index variance
ι_q, k_q	Standardized selection intensity and variance reduction coefficient for category q
ρ, L	Index accuracy and generation interval
$\boldsymbol{d}_{pq}$	For refining α: $\boldsymbol{d}_{pq} = E[\boldsymbol{s}_{j(p)} \mid j(p)$ has category q parent$] - \bar{\boldsymbol{s}}_p$; has dimension $n_s \times 1$; the $n_c n_s \times n_c$ matrix $\boldsymbol{D}$ has elements $\boldsymbol{d}_{pq}$
$b_q(p, t)$, $c_q(p, t)$	For ancestor i in category q at time 0, the genetic contribution to selected individuals in category p time t is $c_q(p, t) + b_q(p, t)(A_{i(q)} - \bar{A}_q)$, with vector of coefficients for all categories p denoted by $\boldsymbol{c}_q(t)$ and $\boldsymbol{b}_q(t)$
$f_q(p, t)$	Regression of $A_{j(p)}$ for selected $j(p)$ at time t on $A_{i(q)}$ for ancestor $i(q)$ at time 0, with vector for all categories p denoted by $\boldsymbol{f}_q(t)$
τ_m, τ_f, τ_w	For sib-indices: regression of the index on the sire's and dam's transmitting ability and on the candidate's Mendelian sampling term;
τ, ι, k	For sib-indices: $\tau = \frac{1}{2}(\tau_m + \tau_f)$; $\iota = \frac{1}{2}(\iota_m + \iota_f)$; $k = \frac{1}{2}(k_m + k_f)$
z, κ	For sib-indices: $z = \rho\sigma_A$; $\kappa = [k\tau + \frac{1}{8}(\tau_m - \tau_f)(k_m - k_f]$

tions was introduced by James and McBride (1958) and was developed by Wray and Thompson (1990) for the prediction of rates of inbreeding (ΔF). Given the fundamental nature of the concept of this article, the definition is restated. The genetic contribution of an ancestor i born at time u to an individual j born at time t ($>u$) is the proportion of the genes of j that are expected to derive by descent from ancestor i. This is different from the definition used by Wray and Thompson (1990), who multiplied this proportion by $X_m + X_f$ (where X_m and X_f are the number of male and female parents in a generation); but as shown by Woolliams *et al.* (1993), a contribution is more usefully defined without this rescaling. It is also distinct from the numerator genetic relationship that considers shared genes, not only those restricted to descent: so full-sibs make no genetic contribution to each other although they have a genetic relationship >0.

The notation is defined to allow extensions to overlapping generations. Therefore contributions are defined within and between categories, where the categories are defined by both age and sex and, potentially, breeding use (*e.g.*, nucleus females and other females). Over its lifetime an individual moves through various categories. An initial objective is to show the relationship between contributions and rate of gain, and for this there is no need to identify details of the category of an individual and what is happening to the different categories over time. For this objective it is necessary only to consider the observed contribution by whatever means it is achieved. However, to develop the concept of gene flow, which is important for understanding the dynamics of overlapping generations, the tracking of categories is required. Therefore, to keep notation minimal at any given stage, the notation for contributions is developed through the article, and a balance between consistency and simplicity was attempted.

The following notation is used initially: $r_{i,u}(j, t)$ is the contribution of ancestor i that was born at time u to individual j born at time t; $r_{i,u}(t)$ is the mean contribution over all the newborn cohort at time t (*i.e.*, one-half of the mean for newborn males plus one-half of the mean for newborn females). For the long-term contributions of i, $r_{i,u} = r_{i,u}(t)$ as $t \to \infty$. For long-term contributions there is less need to specify u, and r_i is used. T_m males and T_f females are scored in each cohort at random, and only scored individuals are candidates for breeding opportunities.

The populations are assumed to mix over time. With mixing, the contribution a particular ancestor makes to later-born individuals tends to a value that is the same for all individuals in later cohorts; *i.e.*, for each i, the variance of $r_{i,u}(j, t)$ among j tends to 0 as $t \to \infty$ (Wray and Thompson 1990). This value is the long-term contribution r_i and will differ between individual ancestors, depending upon the lifetime breeding use of i, its breeding value, and other selective advantages both genetic and nongenetic, and chance factors. Wray and

Thompson (1990) and Grundy *et al.* (1998) describe in more detail the relationship between the long-term contribution and the numerator relationship matrix.

The full development presented in this article assumes the infinitesimal model with negligible rates of inbreeding, because this satisfies the principal requirement for a period of equilibrium in the population structure. This study uses Mendelian sampling terms to mean the deviation of the breeding value of an individual from the mean of its parents' breeding values and Mendelian sampling variance to mean the variance of these deviations.

Rates of gain: The breeding value of an individual may be decomposed into a sum of independent terms involving the breeding values of the base generation and Mendelian sampling terms of all other ancestors. This may be done by observing that (i) the breeding value of an individual j born at time t can be expressed as the average of its parental breeding values plus a deviation (the Mendelian sampling term), which is independent of its parental breeding values, *i.e.*, $A_{j,t} = \frac{1}{2}A_{\text{sire}} + \frac{1}{2}A_{\text{dam}} + a_{j,t}$; and (ii) by going backward through the pedigrees, this substitution can be repeated for each generation of ancestors until the base generation is reached. The coefficients for these terms are the genetic contributions of the ancestors to individual j born at time t. Therefore,

$$A_{j,t} = \sum_{u=1}^{t}\sum_{i} r_{i,u}(j,\, t)a_{i,u} + \sum_{i} r_{i,0}(j,\, t)A_{i,0}.$$

The second term is to allow for the base population, not necessarily unselected, where it is assumed that parents are unknown and so all the genetic information prior to $t = 0$ is contained in this base information. Let G_t, the genetic merit of the population at time t, be the average of the breeding values of the newborn males and females, *i.e.*, $G_t = \frac{1}{2}\Sigma_{j\text{males}} T_{\text{m}}^{-1}A_{j,t} + \frac{1}{2}\Sigma_{j\text{females}}T_{\text{f}}^{-1}A_{j,t}$; then $G_t = \Sigma_{u=1}^{t}\,\Sigma_i\, r_{i,u}(t)a_{i,u} + \Sigma_i r_{i,0}(t)A_{i,0}$. Because $E[a_{i,u}] = 0$, the cross-product $r_i a_i$ is related to the covariance between r_i and a_i; thus sustained genetic gain is related to the creation of covariance between contributions and Mendelian sampling terms.

Let the gain made by selection in cohort t be $\Delta G_t = G_{t+1} - G_t$, and $\Delta r_{i,u}(t) = r_{i,u}(t + 1) - r_{i,u}(t)$; then

$$\Delta G_t = \sum_{u=1}^{t}\sum_{i}\Delta r_{i,u}(t)a_{i,u} + \sum_{i}\Delta r_{i,0}(t)A_{i,0}. \tag{1}$$

Because the population is assumed to mix, the terms $\Delta r_{i,u}(t) \to 0$ as $t \to \infty$ and so $\Delta r_{i,u}(t)a_{i,u} \to 0$ as $t \to \infty$ for a fixed u, and, in particular, the terms for the base population terms in Equation 1 tend to 0. Therefore for large t, summing over males ($i(m)$) and females ($i(f)$) separately and taking expectations,

$$E[\Delta G_t] = \sum_{u=1}^{t}\sum_{q=mf} T_q E[\Delta r_{i(q),u}(t)a_{i(q),u}]. \tag{2}$$

If an equilibrium is approached (as will be the case with the infinitesimal model when inbreeding is ignored), the expected *change* in covariance between r_i and a_i will depend only on $t - u$ and not on u *per se*, *i.e.*, only on the elapsed time since the ancestor's birth, and not on the actual time of birth. So $E[\Delta r_{i(q),u}(t)a_{i(q),u}] = E[\Delta r_{i(q),u+\delta t}(t + \delta t)a_{i(q),u+\delta t}]$.

After making these substitutions, ΔG_t may be expressed as a sum of changes in contributions of individual ancestors, *i.e.*,

$$\sum_{u=1}^{t} E[\Delta r_{i(q),u}(t)a_{i(q),u}] = \sum_{t=0}^{u-1} E[\Delta r_{i(q),u}(u + t)a_{i(q),u}].$$

For u large enough, the right-hand side will approach its equilibrium value $E[r_{i(q),u}a_{i(q),u}]$. Therefore, for a sufficiently large t, $E[\Delta G_t] = E[\Delta G_{\text{eq}}]$ and substitution of these results into Equation 2 gives

$$E[\Delta G_{\text{eq}}] = T_{\text{m}}E[r_{i(m)}a_{i(m)}] + T_{\text{f}}E[r_{i(f)}a_{i(f)}] \tag{3}$$

or equivalently, $E[\Delta G_{\text{eq}}] = T_{\text{m}}\text{cov}(r_{i(m)}, a_{i(m)}) + T_{\text{f}}\text{cov}(r_{i(f)}, a_{i(f)})$. An equivalent expression to Equation 3 can be given as a continuous function of time (available from the authors).

Comparison of Equation 3 with other expressions of gain: The traditional formula for quantitative genetic gain expresses gain as the product of selection intensity (ι), accuracy (ρ), and genetic standard deviation (σ_{A}) defined in a single generation. Equation 3 makes explicit and clear that (i) genetic gain must arise from "good" ancestors contributing more genes; (ii) this process of contributing genes concerns more than a single generation; (iii) sustained gain depends on utilizing the new variation, *i.e.*, the Mendelian sampling variation, entering the population each generation; and (iv) quantitatively, the covariance of r_i with a_i gives a complete description of the process involved in items (i)–(iii).

The traditional expression for gain may be the most tractable form for calculation in most schemes, but it is unclear that this will always be the case, *e.g.*, with quadratic indices as described by Meuwissen (1997) and Grundy *et al.* (1998). However, it is shown that formulae developed in the next sections and used in Equation 3 lead to estimates for rates of gain that are precisely equivalent to the traditional expression for important cases. Therefore, the main outcome of Equation 3 is that the rate of gain has been connected to the pedigree, which is not apparent with $\iota\rho\sigma_{\text{A}}$. Equation 3 is useful for decomposing achieved gain, but its usefulness for prediction is limited because r_i is observed. Therefore, it is necessary to develop expectations for r_i.

Framework for general solution: As described above, one reason for deriving expected long-term contributions is to exploit the relationships between the long-term contributions and rates of gain by replacing the observed r_i. There are other reasons that are perhaps more important: first, the expected contributions are involved in predicting rates of inbreeding (ΔF) in selected populations using the relationship between ΔF and the sum of squared contributions (Wray and Thompson 1990; Woolliams *et al.* 1993); second, the expected long-term contributions represent the expected gene flow in the population, and in complex population structures (with overlapping generations and breeding pyramids) this information is essential for scheme design. To develop expected contributions it is necessary to modify slightly the notation used. In particular, it is necessary for breeding categories (*i.e.*, ages and sexes) to be explicit, so $i(q)$ denotes an ancestor in category q.

The expected long-term contribution of individual $i(q)$ is defined conditional on a vector of n_{s} selective advantages, $\boldsymbol{s}_{i(q)}$. The $\boldsymbol{s}_{i(q)}$ are expressed as deviations from the average of the selected contemporaries $\bar{\boldsymbol{s}}_q$. The selective advantages influence the success of the offspring and (or) may influence the selection of subsequent descendants, *i.e.*, $\mu_{i(q)} = E[r_{i(q)}|\boldsymbol{s}_{i(q)}]$. For example, an expected breeding value (EBV) of an ancestor at the time of selection of its own offspring will influence the number of offspring that are selected and will play a role in the number of grand-offspring selected; in contrast, the corresponding prediction error of the EBV will not influence selection of offspring but will influence selection of grand-offspring. The conditional expectation expresses the expected contribution as a function of the selective advantages. If a linear model for the conditional expectation is assumed, then $\mu_{i(q)} = \alpha_q + \boldsymbol{\beta}_q^T(\boldsymbol{s}_{i(q)} - \bar{\boldsymbol{s}}_q)$. If an equilibrium is assumed, then the coefficients α_q and $\boldsymbol{\beta}_q$ will not change over generations and the same coefficients can be used for both the ancestor and the selected offspring. The expected lifetime long-term contribution of an individual i is the sum of the expected long-term contributions for all categories that i belonged to over its lifetime.

The objective of the following section is to define a set of achievable steps that can be followed to derive formulae for α_q and β_q to obtain expected contributions even in complex breeding schemes. The starting point is to note that the long-term contribution of individual i is given by

$$r_i = \tfrac{1}{2} \sum_{\text{offspring } j} r_j, \tag{4}$$

where the sums are taken over its male and female offspring. Because unselected offspring have no long-term contribution, these sums may be restricted to the selected offspring. Taking expectations conditional on $\boldsymbol{s}_{i(q)}$ and summing over categories p,

$$\mu_{i(q)} = \tfrac{1}{2} \sum_{\text{categories } p} \times E[\text{number offspring selected in category } p|\boldsymbol{s}_{i(q)}] \times E[r_{j(p)}|\boldsymbol{s}_{i(q)}]. \tag{5}$$

Let the population have n_c categories that describe sex, age, and breeding purpose. Discrete generations are a special case with only two categories, males and females. Initially, $\boldsymbol{s}_{i(q)}$ is assumed to be a single variable ($n_s = 1$), namely the breeding value $A_{i(q)}$. This was assumed for mass and sib-index selection by Woolliams *et al.* (1993) and Wray *et al.* (1994). In this situation β_q is a single number. The expected long-term contributions for individual i in category q can then be represented by $\mu_{i(q)} = \alpha_q + \beta_q(A_{i(q)} - \bar{A}_q)$.

The solutions are obtained from four steps: (i) for overlapping generations only, to determine the gene flow from the parents (*sic*) in previous periods to selected individuals in the current period; (ii) to regress the expected number of offspring selected for a parent upon the selective advantage(s), with the regression coefficients λ_{pq} forming an $n_c \times n_c$ matrix Λ; (iii) to regress the selective advantage(s) of a selected offspring upon those of the parent, with the coefficients π_{pq} forming an $n_c \times n_c$ matrix Π; (iv) from these steps calculate the vectors of α_q and β_q for all categories, *i.e*, $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_{n_c})^T$, and $\beta = (\beta_1, \beta_2, \ldots, \beta_{n_c})^T$, both of dimension $n_c \times 1$.

Step 1, defining the gene flow matrix $\boldsymbol{G}$*:* The concept of gene flow (Hill 1974) is used, but the development of Hill does not account for the inheritance of selective advantage that is critical for selection. A consequence of this selective advantage is that the probability that the parent of a selected individual in category p comes from category q will depend on the selection intensity in category p and the selective advantage of category q over other categories contributing candidates for category p. If category q has a selective advantage over other categories then its offspring will have increasing success as selection becomes more intense. Consider an example where dams from age 1 have a higher genetic merit than those of age 2, and the two ages contribute equally to a group of newborn individuals. If selection among this newborn group is at random, then those chosen are expected to come equally from 1- and 2-yr-old females; but if there is selection in this group, offspring of females of age 1 would be expected to be favored.

In the standard gene flow matrix (Hill 1974), the key elements are $g_{0,pq}$ representing the proportions of genes in the newborn cohort from which category p will be selected (at some time in their life) that arise from category q parents. To obtain the expected long-term contributions a modified matrix is required ($\boldsymbol{G}$, of dimension $n_c \times n_c$) in which each row represents a category of *selected* individuals (rather than newborn), and with the elements g_{pq} of each row representing the proportions of genes in the *selected* individuals transferred through breeding from the parents in the different categories q. With discrete generations and the standard two pathways, $\boldsymbol{G} = (\tfrac{1}{2}, \tfrac{1}{2}|\tfrac{1}{2}, \tfrac{1}{2})$ always. Deterministic procedures to obtain $\boldsymbol{G}$ are described in detail by Bijma and Woolliams (1999) and briefly in the application concerning overlapping generations in this article.

Step 2, defining and deriving Λ*:* A regression model is required for the expected number of offspring (the expected selection score) of a parent in category q that are selected to breed in category p on the breeding value of their parent. With random selection the proportion of the X_p selected in category p that are expected to have category q parents is $2g_{0,pq}$ and these are divided equally among the X_q parents in category q. In this case, the expected selection score for a parent in category q is simply a constant $2X_p g_{0,pq} X_q^{-1}$ and does not depend upon $A_{i(q)}$. With selection, appendix a shows that this expectation is of the form $2X_p g_{pq} X_q^{-1}(1 + \lambda_{pq}(A_{i(q)} - \bar{A}_q))$. The elements λ_{pq} form an $n_c \times n_c$ matrix Λ. For mass selection the $\lambda_{pq} = \tfrac{1}{2}\iota_p \sigma_P^{-1}$, where ι_p is the intensity of selection in category p, and σ_P is the phenotypic standard deviation.

Step 3, defining and deriving Π*:* A second regression model is required for the regression of the breeding value of the *selected* offspring on the breeding value of the parent. In principle these, too, depend on both the category of offspring and parent, giving an $n_c \times n_c$ matrix Π, with π_{pq} representing the coefficient for offspring category p and parent category q. Thus $E[A_{j(p)} - \bar{A}_p] = \pi_{pq}(A_{i(q)} - \bar{A}_q)$. appendix b gives a general derivation for Π that is used in all the applications. For the case of mass selection with only the breeding value conferring selective advantage, $\pi_{pq} = \tfrac{1}{2}(1 - k_p h^2)$, where k_p is the variance reduction coefficient for selection in category p and h^2 is the heritability in the candidates.

Step 4, solutions: Using Equation 5 with (i) the breeding value replacing $\boldsymbol{s}_{i(q)}$ as the selective advantage; (ii) the E[number selected offspring$|A_{i(q)}$] replaced by $2X_p g_{pq} X_q^{-1}(1 + \lambda_{pq}(A_{i(q)} - \bar{A}_q))$; (iii) the assumption of equilibrium justifying the use of the same α and β for both parent and offspring; (iv) $(A_{j(p)} - \bar{A}_p)$ in $E[r_{j(p)}|\mathbf{s}_{i(q)}]$ replaced by $\pi_{pq}(A_{i(q)} - \bar{A}_q)$; and collecting terms independent of $A_{i(q)}$ and those linearly dependent upon $A_{i(q)}$ separately gives

$$\alpha_q = \sum_p X_p g_{pq} X_q^{-1} \alpha_p \tag{6a}$$

$$\beta_q(A_{i(q)} - \bar{A}_q) = \sum_p (X_p g_{pq} X_q^{-1} \lambda_{pq} \alpha_p + X_p g_{pq} X_q^{-1} \beta_p \pi_{pq}) \times (A_{i(q)} - \bar{A}_q). \tag{6b}$$

The quadratic terms have been neglected and this is addressed in the discussion. If $\boldsymbol{N}$ is the diagonal matrix with elements X_p, then the matrix forms of Equations 6a and 6b are

$$(\boldsymbol{N}\alpha) = \boldsymbol{G}^T(\boldsymbol{N}\alpha) \tag{7a}$$

$$(\boldsymbol{N}\beta) = (\boldsymbol{I} - \boldsymbol{G}^T \otimes \Pi^T)^{-1}(\boldsymbol{G}^T \otimes \Lambda^T)(\boldsymbol{N}\alpha), \tag{7b}$$

where $\otimes$ denotes element-by-element multiplication of the matrices.

Therefore, $\boldsymbol{N}\alpha$ is a right eigenvector of $\boldsymbol{G}^T$ with eigenvalue 1 (this eigenvector exists because all rows of $\boldsymbol{G}$ sum to 1). This defines α only up to a scalar. Let L be the generation interval defined as the period of time for the population to renew itself. Then (i) over its lifetime, a single cohort has a total long-term contribution of $\Sigma_p X_p \alpha_p$ and so $L\Sigma_p X_p \alpha_p = 1$; (ii) the average age at which the long-term contributions are made is given by $L = (\Sigma X_p \alpha_p)^{-1} \Sigma X_p \alpha_p \text{age}(p)$, where age($p$) is the age of individuals in category p. Combining these two formulae gives the constraint $\Sigma_p X_p \alpha_p \text{age}(p) = 1$, and this is sufficient to define α uniquely. Note $L = (\Sigma X_p \alpha_p)^{-1}$. For discrete generations, with the standard two pathways, $\alpha = (\tfrac{1}{2}X_m^{-1}, \tfrac{1}{2}X_f^{-1})^T$ and $L = 1$ always.

The vector $\boldsymbol{N}\beta$ is completely determined once $\boldsymbol{G}$, Π, Λ, and α are defined. If we consider a simple case with a single category that may occur with a monoecious population with

X parents, then all the terms become scalars and $\beta = (1 - \pi)^{-1}\lambda\alpha$ and $\alpha = X^{-1}$. For more than one category the g_{pq} act as weighting factors across the categories for the different values of π_{pq} and λ_{pq}.

Extension to multiple variables (s): With multiple variables (n_s) conferring selective advantage, $\mu_{i(q)} = \alpha_q + \beta_q^T(\boldsymbol{s}_{i(q)} - \bar{\boldsymbol{s}}_q)$. α remains a vector of length n_c but β is a vector of length $n_c n_s$ of the form $(\beta_1^T, \beta_2^T, \ldots, \beta_{n_c}^T)^T$. Each element λ_{pq} becomes a $1 \times n_s$ submatrix λ_{pq}, and each element π_{pq} becomes an $n_s \times n_s$ submatrix π_{pq}. The matrix Λ is of order $n_c \times n_c n_s$, and Π is of $n_c n_s \times n_c n_s$. The solution for α remains unchanged (Equations 6a and 7a). To obtain the equation analogous to (6b), let $s_{j(p(v))}$ and $s_{i(q(w))}$ represent variables v and w in $\boldsymbol{s}_{j(p)}$ and $\boldsymbol{s}_{i(q)}$, respectively, so $1 \le v, w \le n_s$:

$$\beta_{q(w)}(s_{i(q(w))} - \bar{s}_{q(w)}) = \Big(\sum_{p=1}^{n_c} X_p g_{pq} X_q^{-1} \lambda_{pq(w)} \alpha_p + \sum_{p=1}^{n_c} X_p g_{pq} X_q^{-1} \sum_{v=1}^{n_s} \beta_{p(v)}\, \pi_{p(v)q(w)}\Big) \times (s_{i(q(w))} - \bar{s}_{q(w)}). \quad (8)$$

The matrix forms of the equations for multiple variables in $\boldsymbol{s}_{i(q)}$ (not shown) are the same as in (7), but with (i) the definition of $\otimes$ being extended to mean the multiplication of the submatrices π_{pq} and λ_{pq} by the element g_{pq}; and (ii) in (7b), $\boldsymbol{N}\beta$ is replaced by $\boldsymbol{N} \otimes \beta$; *i.e.*, each subvector β_q is multiplied by X_q.

A further refinement of α: This section is not essential to the overall development, but it can prove important for good approximation in complex structures and it is used in results. The section describes an improvement in the estimation of α, which corresponds to a second-order approximation.

The g_{pq} account for the different selective advantages among the categories of the parents at the time of selection but the advantages or disadvantages are inherited in part by the selected offspring. From Equation 6a, $\alpha_q = \Sigma_{p=1}^{n_c} X_p g_{pq} X_q^{-1}$ $(\alpha_p + \beta_p^T \boldsymbol{d}_{pq})$, where $\boldsymbol{d}_{pq} = E[\boldsymbol{s}_p|\text{category } q \text{ parent}] - \bar{\boldsymbol{s}}_p$. After rearranging terms in Equations 6a and 6b,

$$(\boldsymbol{N}\alpha) = (\boldsymbol{G}^T + (\boldsymbol{G}^T \otimes \boldsymbol{D}^T)(\boldsymbol{I} - \boldsymbol{G}^T \otimes \Pi^T)^{-1} \times (\boldsymbol{G}^T \otimes \Lambda^T))(\boldsymbol{N}\alpha), \quad (9)$$

where $\boldsymbol{D}$ is dimension $(n_c n_s \times n_c)$, with submatrix pq equal to $\boldsymbol{d}_{pq}$. Although α is still defined as a right eigenvector of a matrix with eigenvalue one, the matrix is now more complex. The constraint to define α uniquely is unchanged. When generations are discrete and with the standard two-pathway model, $\boldsymbol{D} = 0$.

Expected long-term contributions and rates of gain: For any one individual i the total long-term contribution is the sum of its long-term contributions as it moves through the different categories over its lifetime, *i.e.*, $r_i = \Sigma_{q=1}^{n_c} r_{i(q)}$. Define $S_{i(q)} = 1$ if i is selected in category q, 0 otherwise; then

$$E[r_i|S_{i(q)}, q = 1, \ldots, n_c] = \sum_{q=1}^{n_c} S_{i(q)} E[r_{i(q)}|S_{i(q)} = 1] = \sum_{q=1}^{n_c} S_{i(q)} \mu_{i(q)}.$$

When the expected long-term contribution is expressed in terms of the components of the breeding value, in particular the Mendelian sampling term, the expected long-term contribution is sufficient for the prediction of genetic gain because the remaining part $(r_{i(q)} - \mu_{i(q)})$ has no covariance with the Mendelian sampling term. Within a category q the sum of $S_{i(q)}$ over all candidates is X_q, and so application of Equation 3 gives

$$E[\Delta G_{\text{eq}}] = \sum_{q=1}^{n_c} X_q E[\mu_{i(q)} a_{i(q)}], \quad (10)$$

where now the expectations are conditional on being selected as a parent rather than unconditional as was the case in Equation 3. Equation 10 is expressed solely in terms of the selected individuals and in terms that are predictable rather than simply observed.

If $\mu_{i(q)} = \alpha_q + \beta_q^T(\boldsymbol{s}_{i(q)} - \bar{\boldsymbol{s}}_q)$ then Equation 10 immediately decomposes the gain into two components: the first, $\Sigma_q^{n_c} X_q \alpha_q E[a_{i(q)}]$, is the expected gain from selection within families, which occurs at the time of selection of the ancestor, while the second, $\Sigma_{q=1}^{n_c} X_q \beta_q^T E[(\boldsymbol{s}_{i(q)} - \bar{\boldsymbol{s}}_q) a_{i(q)}]$, represents the expected between-family gain, and describes the changes in contribution of selected ancestors from the time of their selection until convergence in the long term. Because the between-family gain is explicitly defined in terms of the selective advantages, the gain can be decomposed into components arising from each category and each selective advantage within categories.

The covariance between the Mendelian sampling term $a_{i(q)}$ and $(\boldsymbol{s}_{i(q)} - \bar{\boldsymbol{s}}_q)$ following the selection of the ancestor can be calculated using standard index theory. Note that because this is a covariance with the deviation from a sample mean, adjustments of $(1 - X_q^{-1})$ should result in increased precision. For simplicity, this has *not* been applied in the results presented. The predicted increase in precision can be confirmed from the results shown.

Development of contributions over time: This section is not essential to the overall development but describes the solution to an important application of gene flow. In complex population structures it is often useful to predict how quickly improvement in one part of the population diffuses through to other parts of the population or what proportion of the gene flow arises from particular pathways (*e.g.*, by male descent alone). This requires methods to predict the rate of convergence of genetic contributions over time.

To simplify the notation the development of contributions over time is given for the single selective advantage, the breeding value, A. It is assumed that when $t = 0$, the population is already in equilibrium. For category q, a *selected* individual at time 0 has a vector (dimension $n_c \times 1$) of contributions to *selected* individuals in category p at time t given by $c_q(p, t) + b_q(p, t)(A_{i(q)} - \bar{A}_q)$. This is a form similar to that of the long-term contribution, but before convergence it will differ between categories p and so needs to be defined for each p. Let $\boldsymbol{c}_q(t) = (c_q(1, t), c_q(2, t), \ldots c_q(n_c, t))^T$, and $\boldsymbol{b}_q(t) = (b_q(1, t), b_q(2, t), \ldots b_q(n_c, t))^T$. Then $\boldsymbol{c}_q(0) = 0$ except for X_q^{-1} in the qth position, and $\boldsymbol{b}_q(0) = 0$. A further vector of regressions is required, $\boldsymbol{f}_q(t)$, for which the pth element is the regression of the breeding value of the selected individual in category p at time t on the breeding value of an ancestor in category q. By definition $\boldsymbol{f}_q(0) = 0$ except for the qth position where it is 1.

It is critical to note that the contributions at time t to the selected individuals in category p of age(p) will depend on the consequences of the selection upon the parental gene pool at time $t -$ age(p): the more intense the selection, the more those parent categories with greater selection advantages will dominate. In a selection scheme, a group of newborn individuals will typically be subject to different selection intensities as they become older. Therefore the complete spectrum of contributions among the selected individuals in the different categories at time t will depend on states back to $t -$ maxage, where maxage is the maximum age of the parents in the breeding scheme. Define $\boldsymbol{G}_p$ to be the $n_c \times n_c$ matrix consisting of zeros, except for the single row corresponding to category p, which is identical to the pth row of $\boldsymbol{G}$. Then

$$\boldsymbol{c}_q(t) = \sum_{p=1}^{n_c} \boldsymbol{G}_p \boldsymbol{c}_q(t - \text{age}(p)) \quad (11a)$$

$$\boldsymbol{b}_q(t) = \sum_{p=1}^{n_c} \boldsymbol{G}_p \boldsymbol{b}_q(t - \text{age}(p)) + \sum_{p=1}^{n_c} X_q^{-1}(\boldsymbol{G}_p \otimes \Lambda) \times f_q(t - \text{age}(p)) \quad (11b)$$

$$f_q(t) = \sum_{p=1}^{n_c} (\boldsymbol{G}_p \otimes \Pi) f_q(t - \text{age}(p)). \quad (11c)$$

Equation 11a describes the contribution of category q to each category at each time t, with element p of the sum describing the contributions of category q ancestors (at time $t = 0$) to category p parents at time t, accounting for the selection in category p through the matrix $\boldsymbol{G}_p$. Equation 11b describes the relationship of contributions from ancestors *within* category q (at time $t = 0$) to each category at each time t to the selective advantage; this arises from two processes, the first, analogous to (11a), from the transfer of differential contributions among ancestors of category q that were accumulated up to and including time $t - 1$, and the second from further differential contributions from selective advantages among the candidates at time t due to ancestors in category q at time $t = 0$. Equation 11c describes the changes in the selective advantages among the candidates at time t due to ancestors of category q at time $t = 0$.

When t becomes large, the mixing assumption for the population ensures that both $\boldsymbol{c}_q(t)$ and $\boldsymbol{b}_q(t)$ converge to a vector with all elements equal, namely $\alpha_q\mathbf{1}$ and $\beta_q\mathbf{1}$, respectively, where $\mathbf{1} = (1, \ldots, 1)^T$. Furthermore $f_q(t) \to 0$ because the eigenvalues of $\boldsymbol{G} \otimes \Pi$ are <1 and >-1, and this reflects the diminishing effect of ancestors over time on the selection advantage of their descendants.

By redefining the state vector at time t to include not only $\boldsymbol{c}_q(t)$ but also $\boldsymbol{c}_q(t-1), \ldots \boldsymbol{c}_q(t - \text{maxage} + 1)$, Equation 11a can be reformulated (results not shown) so that the state vector at time t is the product of a square stochastic matrix of order $n_c \times$ maxage and the state vector at time $t - 1$. Using this reformulation and the properties of stochastic matrices (described in Appendix 1 of Hill 1974), it can be demonstrated that Equations 11 are consistent with Equations 7 and the constraint $\Sigma_p X_p \alpha_p \text{age}(p) = 1$ (results not shown).

The discrete time contributions with the refinement in estimating α are given in appendix c. An example of application is given in results.

APPLICATION OF MODELS AND RESULTS

Expected long-term contributions and genetic gain for general sib-indices in discrete generations: A general sib-index of the form $I = b_1(P - \bar{P}_F) + b_2(\bar{P}_F - \bar{P}_H) + b_3\bar{P}_H$ was studied by Wray *et al.* (1994), where I is the index, P is the phenotype of the candidate, $\bar{P}_F$ is the mean of the full-sib family (size n_F) including the candidate, and $\bar{P}_H$ is the mean of the half-sib family (size n_H) including the candidate and full-sibs. Mass selection is a special case with $b_1 = b_2 = b_3 = 1$. For simplicity, the only selective advantage considered in this article, $\boldsymbol{s}_{i(q)}$, is the breeding value $A_{i(q)}$, with other forms of environmental influences that are often considered (*e.g.*, litter effects) omitted and random mating assumed. For discrete generations there are just two categories, males and females. In an unselected base generation the phenotypic variance (σ_P^2) is 1 and the additive genetic variance is h_0^2. The categories are $q = m$ for male and f for female. The notation is included in Table 1.

The regression models required are derived from appendices a and b: $\lambda_{pq} = \iota_p\tau_q(2\sigma_I)^{-1}$ and $\pi_{pq} = \frac{1}{2}(1 - k_p\tau_q\rho\sigma_A\sigma_I^{-1})$, where $\tau_m = b_3$ and $\tau_f = b_2(1 - X_mX_f^{-1}) + b_3X_mX_f^{-1}$ and $\tau = \frac{1}{2}(\tau_m + \tau_f)$. The τ_q values were used by Wray *et al.* (1994) and are twice the regression of the index of the candidate on the breeding value of the parent of sex q, σ_I^2 is the variance of the index, and ρ is the accuracy of the index.

After simplification of Equation 7 (see Woolliams *et al.* 1999 for further details),

$$\alpha_q = \tfrac{1}{2}X_q^{-1}, \quad \beta_q = \tfrac{1}{4}\iota(\tau + \tau_q)(\sigma_I + \kappa z)^{-1}X_q^{-1}, \quad (12)$$

where $\kappa = [k\tau + \frac{1}{8}(\tau_m - \tau_f)(k_m - k_f)]$ and $z = \rho\sigma_A$. This form is nearly equivalent to that given by Wray *et al.* (1994), but their derivation proceeded on different (and more complex) lines. Three points of difference should be noted. First, Wray *et al.* (1994) do not include the small $\frac{1}{8}(\tau_m - \tau_f)(k_m - k_f)$ term in κ that arises when *both* the selection intensity and the regression on the parental breeding value differ between the sexes. Second, the indices of Wray *et al.* (1994) are explicitly scaled so that the regression of the breeding value of the candidate on its index is 1 (*i.e.*, $\rho\sigma_A\sigma_I^{-1} = 1$), but scaling does not change $\tau_q\sigma_I^{-1}$ and so α and β do not change with scaling. Finally in this article, predictions in equilibrium are obtained using equilibrium parameters.

Rate of gain from sib-indices: The decomposition of the rate of gain is achieved using Equation 10 and standard index theory. Within-family gain is given by

$$\sum_q^{n_c} X_q\alpha_q E[a_{i(q)}|i \text{ selected}] = \sum_{q=m,f} \tfrac{1}{4}h_0^2\iota_q\tau_w\sigma_I^{-1} = \tfrac{1}{2}h_0^2\iota\tau_w\sigma_I^{-1},$$

because $\alpha_q = \frac{1}{2}X_q^{-1}$ and $E[a_{i(q)}|i \text{ selected}] = \frac{1}{2}h_0^2\tau_q\tau_w\sigma_I^{-1}$, where τ_w is the regression of the index I on $a_{i(q)}$ ($\tau_w = b_1(1 - n_F^{-1}) + b_2(n_F^{-1} - n_H^{-1}) + b_3n_H^{-1}$). The total between-family gain is given by

$$\sum_{q=1}^{n_c} X_q\beta_q^T E[(\boldsymbol{s}_{i(q)} - \bar{\boldsymbol{s}}_q)a_{i(q)}|i \text{ selected}] = \sum_{q=m,f} \tfrac{1}{8}h_0^2\iota(\tau + \tau_q)(1 - k_q\tau_w z\sigma_I^{-1})(\sigma_I + \kappa z)^{-1}$$

because $\text{cov}(a_{i(q)}, A_{i(q)}) = \frac{1}{2}h_0^2(1 - k_q\tau_w z\sigma_I^{-1})$ for the selected individuals in category q.

The total gain, summed over both sexes, including both between- and within-family gain is, after simplification,

$$\Delta G_{eq} = \tfrac{1}{2}h_0^2\iota(\tau_w + \tau)(\sigma_I + \kappa z)^{-1}. \quad (13)$$

This uses the result $k_m\tau_m + k_f\tau_f = \frac{1}{2}(k_m + k_f)(\tau_m + \tau_f) + \frac{1}{2}(k_m - k_f)(\tau_m - \tau_f) = 2k\tau + \frac{1}{2}(k_m - k_f)(\tau_m - \tau_f)$.

Consistency with other approaches: Equation 13 for equilibrium ΔG_{eq} can be compared to the standard formula $\Delta G = \iota\rho\sigma_A = \iota z$. Equation 13 comes from considering the gain achieved from a single cohort over all subsequent generations, whereas the standard formula comes

from considering the gain achieved by all previous generations over a single cohort. For an equilibrium the two forms must be equal, and equating them results in a quadratic equation for z:

$$\kappa z^2 + \sigma_I z - \tfrac{1}{2}h_0^2[\tau_w + \tau] = 0. \quad (14)$$

Equation 14 can be obtained as an equilibrium condition when using standard index theory with $\sigma_A^2 = \tfrac{1}{2}h_0^2 + \tfrac{1}{4}\sigma_A^2(1 - k_m\rho^2) + \tfrac{1}{4}\sigma_A^2(1 - k_m\rho^2)$ and $\mathrm{cov}(A, I) = \rho\sigma_A\sigma_I$.

This demonstrates a consistency between the methods presented in this article (in particular those detailed in appendices a and b) with results from classical index theory for discrete generations. Thus the decision to neglect the second-order correction for the Bulmer effect when deriving π_{pq} in appendix b (*i.e.*, correcting the genetic variance of the selected parents for selection among their offspring) is also implicit in standard index theory.

Equation 14 can be used to give reasonable estimates of equilibrium gain for indices even when using unselected base parameters, because many of the terms are constant over time. To use Equation 14 only the base generation value of σ_I is required to solve the quadratic equation for z and then gain is estimated by ιz. Using (14) to obtain z results in underestimates rather than the overestimates obtained using base parameters and ignoring linkage disequilibrium. However, the magnitude of the errors from (14) is qualitatively smaller (Woolliams *et al.* 1999). Estimates from (14) are not precise because they assume σ_I constant, and further improvements to Equation 14 would require an iterative scheme in combination with $\sigma_I^2 = \sigma_I^2 - \tfrac{1}{4}z^2([b_2^2(1 - X_mX_f^{-1}) + b_3^2X_mX_f^{-1}](k + k_f) + b_3^2(k + k_m))$. The consistency, demonstrated with standard index theory, shows that this leads to the same result as the usual procedures for deriving equilibrium gain by iterating on the index accuracy and the genetic variance among the parents.

Expected long-term contributions for best linear unbiased predictors: The analysis of individual long-term contributions can be extended to BLUP evaluation and indices derived from it. With sib-indices, $\boldsymbol{s}_{i(q)}$ was simply the breeding value $A_{i(q)}$ because it is the only means by which a parent may influence its offspring over multiple generations (in the absence of common environmental effects, etc.). With BLUP, different approaches to the form of $\boldsymbol{s}_{i(q)}$ can be taken. Woolliams *et al.* (1999) used three terms for individual i in category q: $\hat{A}_{i(q)}$, the "initial EBV" at the point of selection of i; $\delta\hat{A}_{i(q)}$, the "increment" in the EBV at the point of selection of its offspring; and $\hat{e}_{i(q)}$, the remaining "prediction error" of the parent at the selection of offspring. Selection of i itself is determined by $\hat{A}_{i(q)}$, the selection of the offspring is influenced by $\hat{A}_{i(q)}$ and $\delta\hat{A}_{i(q)}$), while selection of grand-offspring and subsequent generations is influenced by all three. Using the methods described here, Woolliams *et al.* (1999) obtained excellent predictions of expected contributions and genetic gain. They showed that the primary source of between-family selection among ancestors in BLUP is the increment in the EBV between its own selection and that of its offspring. The initial EBV played the least important role.

Extensions to other inheritance modes in the absence of allelic interactions: Extensions of the model to other inheritance modes, such as additive maternal effects or X-linked variation, are made by defining the variables in $\boldsymbol{s}_{i(q)}$ and their impact on λ_{pq} and π_{pq}. As an example, results with maternal imprinted variation are given, where the passage of genes from parent to offspring follows normal Mendelian inheritance, but only the alleles passed to the offspring by the dam are expressed and affect the phenotype. For maternal imprinting, the breeding value can be split into the "expressed" breeding value (A^+) inherited from the dam, and the "latent" breeding value (A^-) inherited from the sire and not expressed.

Define $\boldsymbol{s}_{i(q)} = (A_{i(q)}^-, A_{i(q)}^+)$, with discrete generations giving two categories, m for males and f for females. In this case, λ_{pm} will be zero because the genes passed by the sire do not influence selection of its offspring. However, λ_{pf} will depend on both breeding values, because although A^- is not expressed in the dam it is expressed in its offspring. For π_{pq}, there is a dependence on both breeding values: genes passed by the sire only affect A^-, and genes passed by the dam only affect A^+. Because genes passed by the sire are not expressed, the regression of offspring on parent is unaffected by selection. Therefore, applying appendices a and b,

$$\boldsymbol{G} = (\tfrac{1}{2}, \tfrac{1}{2}|\tfrac{1}{2}, \tfrac{1}{2})$$

$$\Lambda = (0.0, 0.0, \tfrac{1}{2}\iota_m\sigma_P^{-1}, \tfrac{1}{2}\iota_m\sigma_P^{-1}|0.0, 0.0, \tfrac{1}{2}\iota_f\sigma_P^{-1}, \tfrac{1}{2}\iota_f\sigma_P^{-1})$$

$$\Pi = (\tfrac{1}{2}, \tfrac{1}{2}, 0.0, 0.0|0.0, 0.0, \tfrac{1}{2}(1 - k_mh^2), \tfrac{1}{2}(1 - k_mh^2)$$
$$|\tfrac{1}{2}, \tfrac{1}{2}, 0.0, 0.0|0.0, 0.0, \tfrac{1}{2}(1 - k_fh^2), \tfrac{1}{2}(1 - k_fh^2)),$$

where $h^2 = \mathrm{Var}(A^+)/\sigma_P^2$, and the phenotypic variance, σ_P^2, is the sum of the variance of A^+ and the environmental variance. Equation 7 was used to obtain β.

Predictions were made using variance parameters obtained after iteration to equilibrium. To calculate ΔG, the expected values of the Mendelian sampling terms for selected individuals and the covariance with $\boldsymbol{s}_{i(q)}$ for selected individuals were calculated using standard index theory:

$$E[(a_m^-, a_m^+, a_f^-, a_f^+)] = (0.0, \tfrac{1}{2}h_0^2\iota_m\sigma_P^{-1}, 0.0, \tfrac{1}{2}h_0^2\iota_f\sigma_P^{-1})$$

$$\mathrm{cov}(\boldsymbol{s}_{i(q)}, (a_{i(q)}^- + a_{i(q)}^+)) = \tfrac{1}{2}h_0^2\,(1.0, (1.0 - k_mh^2), 1.0, (1.0 - k_fh^2))^T.$$

Because this is imprinted variation, half the genes from an ancestor will be expressed in females and half will be latent in males in the long term. Therefore gains predicted from Equation 3 should be halved.

Excellent predictions of expected genetic contribu-

tions and genetic gain were obtained (Woolliams *et al.* 1999). From these results Woolliams *et al.* (1999) were able to show the relative importance of A^- and A^+ in male and female parents in contributing to within- and between-family gain.

Overlapping generations: An example of application with overlapping generations is presented for mass selection, with a fixed number of parents selected at each age, in a two-path scheme (*i.e.*, there was no subdivision of breeding individuals into males to breed males, males to breed females, etc.). The general approach is explained in more detail by Bijma and Woolliams (1999). The steps are illustrated using a scheme with three categories: 20 males breeding at 1 yr of age, 20 females breeding at 1 yr of age, and 20 females breeding at 3 yr of age, respectively. The number of offspring per litter was eight and the trait was assumed to have a heritability of 0.4. The age groups not used for parents are omitted: males age 1 (category 1), females age 1 (category 2), and females age 3 (category 3).

1. The genetic make-up of the newborns is described by $g_{0,p1}$, $g_{0,p2}$, and $g_{0,p3}$. These are 0.5, 0.25, and 0.25, respectively for all categories p. It is the same for all offspring categories p because it is only a two-path model. From the $g_{0,pq}$, and the number of parents and the family sizes, the selection intensities (ι_p) and variance reduction coefficients (k_p) were calculated for each category: $\iota_p = 1.647$, $k_p = 0.817$, *i.e.*, the same for all three categories.
2. An initial ΔG was assumed as a starting point for iteration. In the following, the starting point was ΔG calculated from standard gene flow (Hill 1974). After iterating to an equilibrium, this was calculated to be $\Delta G = 0.412$.
3. The genetic value of the selected parents in category p was $\iota_p h^2 \sigma_P - (\text{age}(p) - 1)\Delta G$. Deviations from the overall means of the selected males and females were $\delta = (0, +0.412, -0.412)$; *i.e.*, the female parents age 1 had breeding values 0.412 units above average and the female parents age 3 had breeding values 0.412 units below average.
4. Before selection, genetic variance in category p was calculated using the pooled variance within categories plus between categories plus the Mendelian sampling variance:

$$\tfrac{1}{2}h_0^2 + \sum_q (\tfrac{1}{4}\sigma_A^2(2g_{0,pq})(1 - k_q h^2) + \tfrac{1}{4}(2g_{0,pq})\delta_q^2).$$

   This was 0.370 for all p, and the phenotypic variance was $\sigma_P^2 = 0.970$ for all p.
5. $\boldsymbol{G}$ was calculated using a truncation algorithm to find a truncation point for a given upper-tail probability for a mixture of Normal distributions. The algorithm was used twice for the selection of candidates in each category, first to obtain the genetic make-up from sire categories and then to obtain the genetic make-up from dam categories. For category p candidates, the mixing proportions for the Normal distributions were $2g_{0,pq}$ ($q = 1, 2, 3$), *i.e.*, the frequency of the candidates with parent category q; the means of the Normal distributions were the deviations of the candidates with parent category q from the mean of all like-sexed candidates, *i.e.*, $\tfrac{1}{2}\delta_q$; and the variances were assumed independent of parent category q and the phenotypic variance was adjusted for the component of genetic variance between categories of the same sex as parent category q, *i.e.*, $\sigma_P^2 - \Sigma_{q^* \text{ same sex as } q} \tfrac{1}{4}(2g_{0,pq^*})\delta_{q^*}^2$. In the first iteration, each row of $\boldsymbol{G}$ was (0.5, 0.336, 0.164), thus indicating that although the dams of ages 1 and 3 provided equal numbers of candidates, the candidates with dams of age 1 were expected to be twice as successful in having selected offspring.
6. Λ and Π matrices were constructed according to appendices a and b, respectively. For mass selection, $\pi_{pq} = 0.5(1 - k_p h^2)$ and $\lambda_{pq} = 0.5\iota p \sigma_P^{-1}$. In the first iteration, $\Pi = 0.344\ \mathbf{1}\mathbf{1}^T$, where $\mathbf{1}^T = (1, 1, 1)$, $\Lambda = 0.836\ \mathbf{1}\mathbf{1}^T$, and $\boldsymbol{D} = \mathbf{1}\ (0, 0.092, -0.188)$. The result for $\boldsymbol{D}$ indicates that the breeding value of a selected individual (of any category p) with a dam of age 1 is expected to be 0.28 greater than a selected individual of the same category with a dam of age 3.
7. α and β were calculated according to Equations 7b and 9. In the first iteration $(\boldsymbol{N}\alpha)^T = (0.395, 0.289, 0.106)$ and $(\boldsymbol{N}\beta)^T = (0.503, 0.338, 0.165)$.
8. The covariance of the Mendelian sampling term with the breeding values was calculated and ΔG was updated using Equation 11; this uses the result that $E[a_{i(q)}] = \tfrac{1}{2}h_0^2 \iota_q \sigma_P^{-1}$, and after selection $\text{cov}(a_{i(q)}, A_{i(q)}) = \tfrac{1}{2}h_0^2(1 - k_q h^2)$.
9. Steps 3 through 8 were repeated to convergence.

Results after convergence of the iterations were $\alpha = (0.0200, 0.0149, 0.0050)^T$ and $\beta = (0.0255, 0.0171, 0.0084)^T$. Predicted gain within families was (0.134, 0.100, 0.034), and predicted gain between families was (0.067, 0.045, 0.022), giving a total gain of 0.402. At equilibrium $\boldsymbol{G}$ was $\mathbf{1}$ (0.500, 0.335, 0.165). This was compared to simulation results for 1000 replicates: $\alpha = (0.0197, 0.0145, 0.0052)^T$ with a maximum SE of 0.0009; $\beta = (0.0249, 0.0175, 0.0071)^T$ with a maximum SE of 0.0004; and a total gain of 0.398 (SE 0.001). Thus very close agreement between simulations and predictions was obtained. As in discrete generations the gain from mass selection was evenly divided between males and females. The gene flow predicted using Hill (1974) is $\alpha = (0.0167, 0.0083, 0.0083)$. Hill (1974) makes no prediction of β.

The generation interval, defined by the time taken to turn over the genes once, was predicted from $(\Sigma X_q \alpha_q)^{-1}$ to be 1.25 (*cf.* 1.26 with SE 0.01 in the simulations), which was notably shorter than the average age of the parents. This was because of the cumulative effect of the selective advantage of the younger age group

TABLE 2

The time course of expected contributions from an individual female parent of age one at $t = 0$

	To males age 1		To females age 1		To females age 3	
Time	$c(t)$	$b(t)$	$c(t)$	$b(t)$	$c(t)$	$b(t)$
$t = 1$	0.0167	0.0140	0.0167	0.0140	0	0
$t = 2$	0.0151	0.0157	0.0151	0.0157	0	0
$t = 3$	0.0132	0.0146	0.0132	0.0146	0.0167	0.0140
$t = 6$	0.0148	0.0168	0.0148	0.0168	0.0145	0.0160
$t = 10$	0.0149	0.0170	0.0149	0.0170	0.0149	0.0170

The breeding scheme has mass selection with 20 male parents of age 1, 40 female parents of ages 1 and 3 (20 of each age), eight offspring per litter, and heritability 0.4. The expected contribution is $c(t) + b(t)(A_i - \bar{A})$.

of females. Although they produced equal numbers of offspring they produced more than twice as many parents. However, the generation interval was not predictable from the equilibrium $\boldsymbol{G}$ alone (*i.e.*, accounting for a single generation of selective advantage) because this would have predicted an interval of 1.33 (*i.e.*, $0.5 \times 1 + 0.335 \times 1 + 3 \times 0.165$).

To obtain the time course of the contributions, appendix c was used. appendix c needs the following matrices based on $\boldsymbol{G}$:

$\boldsymbol{G}_1$: (0.500, 0.335, 0.165|0.0, 0.0, 0.0|0.0, 0.0, 0.0)

$\boldsymbol{G}_2$: (0.0, 0.0, 0.0|0.500, 0.335, 0.165|0.0, 0.0, 0.0)

$\boldsymbol{G}_3$: (0.0, 0.0, 0.0|0.0, 0.0, 0.0|0.500, 0.335, 0.165).

The results are shown in Table 2 for the time course of contributions from category 2. The contributions converged in cohort 10.

DISCUSSION

This study developed a framework for predicting the expected genetic contributions of individuals and categories of individuals under a wide range of selection and inheritance models. This framework allows selection to be more properly accounted for compared to existing gene-flow methods for overlapping generations and multiple breeding groups (such as that presented by Hill 1974). Furthermore, it advances understanding by considering the differential gene flow among individuals within categories, an extension not hitherto achieved except in some special cases. The framework was constructed by first modeling the selection process and the transfer of selective advantages within a single generation of selection, and second, extending this to multiple generations. Two regression models are required, both of which are derived using standard index theory: first, a model describing the expected number of selected offspring a parent may have (Λ); and the second describing the relationship of the selective advantages of a selected offspring with those of the parent (Π). Predictions of genetic gain directly follow from the expected long-term contributions. Unlike $\iota\rho\sigma_A$, the relationship between gain and contributions (Equations 3 and 10) shows that gain comes from generating a covariance between the long-term contributions and the new variance arising in the population (*i.e.*, the Mendelian sampling variation) in each cohort, thus changing the description of gain from a statistical one to a genetical one.

The framework has been developed to describe the expected genetic contribution over all time horizons from the short-term to the long-term. The novel, closed formulae (Equations 7 and 9) produced for the expected long-term contribution of an ancestor rely on the assumption of equilibrium in the selection process. If there is no equilibrium the error will depend on the relative degree of departure in relation to the timescale of convergence of the contributions (approximately five generations). However, this assumption is not necessary for the use of Equations 11, where contributions are predicted over finite time periods, but more effort may be required to define the changes in the necessary parameters if there is no equilibrium.

In the development of the framework, the effects of inbreeding on parameters and progress have been neglected, but this is not a serious problem. First, the timescale for the convergence of contributions is small in comparison to the timescale for the effects of inbreeding on parameters in breeding schemes, especially where inbreeding is controlled to be at reasonable levels. The impact of individuals within a cohort is very largely decided within five generations, and even within this period, the scope for controlling an individual's contribution declines exponentially (the scope can be measured by the variance of an individual's contribution within the population). A second reason is that schemes will most usefully be compared at the same rates of inbreeding, and so the neglect of inbreeding is less likely to bias the comparisons made.

The expected long-term contribution has been described in a general linear form $\alpha_q + \beta_q^T(\boldsymbol{s}_{i(q)} - \bar{\boldsymbol{s}}_q)$,

where $\boldsymbol{s}$ is a vector of selective advantages for an ancestor *i*. Judged by the accuracy of the results in this study, the omission of quadratic terms from the model has not led to serious errors in predicting the rates of gain or the linear component of relationship between the long-term contribution and the selective advantages. Quadratic terms in $\boldsymbol{s}$ do not affect the prediction of rates of gain unless terms of the order $E[\boldsymbol{s}^2a]$ are significant (which will involve the skewness of a after selection), and will not influence the predicted rate of inbreeding unless higher moments than the variance of $\boldsymbol{s}$ are considered (Woolliams and Thompson 1994). The linear approximations used in the applications, and presented in the appendices, were robust.

The α represents the proportion of genes that derive from the various categories as a whole, and these can differ qualitatively from predictions using Hill (1974), because the earlier study does not account for the inheritance of selective advantages. The impact of this may be particularly great where breeding structures, subject to selection, are subdivided with migration, either planned or random, taking place between the subdivisions. In these circumstances, ignoring the selective advantage between groups will overestimate the impact of groups of lesser merit and underestimate the impact of cohorts of greater merit. The consequences of these errors may be the maintenance and use of subdivisions that have little potential to contribute in the long-term and a greater rate of inbreeding in the population than had been anticipated (Bijma *et al.* 1999). The framework presented here and that of Hill (1974) give the same prediction of α when selection is at random, because (i) elements of $\boldsymbol{G}$ are identical to $g_{0,pq}$, (ii) $\Pi = 0$, and (iii) $\Lambda = 0$.

The genetic contribution of an individual represents the expected impact its Mendelian sampling term has on the population. Within a cohort, the magnitude of the contribution made by an individual will depend upon the breeding categories in which it is included over its lifetime. In any newborn cohort, even when generations overlap, the males are expected to have a total long-term contribution equal to those of the females, *i.e.*, $\Sigma_{\text{male categories}} X_q\alpha_q = \Sigma_{\text{female categories}} X_q\alpha_q$. When generations are discrete these sums are equal to one-half, but when generations overlap the sums will be less than one-half.

The sum of the total contributions from any one cohort, including both sexes, is a natural measure of the rate at which genes in the population are renewed. In particular the rate measured by the $\Sigma X_q\alpha_q$ places an emphasis upon those contributions that are destined to remain in the population in the long term. Thus $(\Sigma X_q\alpha_q)^{-1}$ is the period of time for the population to complete a cycle of renewal and is a measure of the generation interval, L. The generation interval defined by the long-term contributions is shorter than the traditional "average age of the parents at the birth of their offspring" for the examples considered, because the younger breeding groups had a selective advantage and the progeny of older parents were less likely to be selected. The need for a modified generation interval arising from the inheritance of the selective advantage was considered previously (Bichard *et al.* 1973; James 1977). Bichard *et al.* (1973) argued that the traditional generation interval might be usefully modified to account for nonrandomness among parental age-groups in the survival of their offspring to produce grand-offspring. This is what occurs with the inheritance of selective advantage between categories of different ages. For example, such a modification would exclude from the calculation of generation interval parents whose sole purpose is to produce a commercial cohort outside the breeding population. James (1977) moved the argument forward by considering the generation interval calculated from only those parents with selected offspring, and he showed that for the purposes of calculating rates of genetic gain either definition of L would suffice providing the calculation of the selection differential is matched to the definition of the generation interval.

The average age of the parents might generally be considered to refer to the age at the birth of unselected offspring. The definition of James (1977) considers the average age of the parents at the birth of the selected offspring who will then produce the unselected *grand-offspring*. These definitions may be viewed as a one-generation estimate of the generation interval and an iteration beyond this, respectively, whereas the calculation from long-term contributions represents the converged estimate. The definition of the generation interval from long-term contributions avoids any debate on what parents should or should not be included. The average age of the parents at the birth of their unselected offspring remains of operational significance to breeding schemes, but the generation interval defined by the long-term contributions is an unambiguous *genetic* property of a population.

The consistency of the framework with other approaches for estimating gain in discrete generations is important, but this consistency does not extend to overlapping generations. The main approach for prediction of gain in overlapping generations is that of Rendel and Robertson (1950). The formula obtained by Rendel and Robertson was also obtained by Hill (1974) as a consequence of deriving the traditional gene flow, and this apparent consistency added credence to both the approach and the wider results of traditional gene flow. However, this study shows that this consistency is not justified. The estimates of equilibrium gain using contributions and Rendel and Robertson differ slightly from each other. The estimate of gain from contributions arises from the prospective analysis of the impact of a single cohort to the future population over the long term. In contrast, the estimate of gain from

Rendel and Robertson (1950) arises from a retrospective analysis of the impact of selection in the whole population to a single cohort. One reason why differences between these approaches might be expected with overlapping generations is the calculation of selection differentials, because each cohort is a mixture of many truncated Normal distributions.

The second component of the expected long-term contribution is the linear regression on the selective advantages of an individual (β). These terms describe the expected differential contributions within a category that will occur during the selection process as a result of the differences in selective advantages. These differential contributions represent the success of one ancestor's descendants over those from another ancestor and therefore measure the expected extent of between-family selection. The between-family selection is responsible for the greater rates of inbreeding that can occur when selection is practiced, and the control of the magnitude of the regression coefficients (and the components of $\boldsymbol{s}$) is an important aspect of methods to optimize the genetic gain with constrained inbreeding rates (*e.g.*, Verrier *et al.* 1993; Villanueva and Woolliams 1997).

The between-family selection may develop very quickly, so that its extent is largely established in the selection of the progeny, or more slowly. This time-course is controlled by $\boldsymbol{G} \otimes \Pi$ and powers of $\boldsymbol{G} \otimes \Pi$, which describe the decay of the ancestor's selective advantage through progeny [see equation for $f_q(t)$ in (11)]. This rate of decay is controlled by the eigenvalues of $\boldsymbol{G} \otimes \Pi$. In the example given for BLUP, the maximum eigenvalue of $\boldsymbol{G} \otimes \Pi$ was 0.18, which may be compared to 0.36 for mass selection with the same numbers of parents and the same initial heritability. Therefore it is clear that a higher proportion of the ultimate between-family selection generated by selection with BLUP is achieved in the first and second generations after the ancestor than is the case with mass selection. This difference has a consequence for the accuracy of the prediction of rates of inbreeding using techniques accounting for coselection in one and two generations (Wray *et al.* 1990) and explains why these methods are notably more accurate with BLUP selection than with mass selection (T. H. E. Meuwissen, personal communication).

The importance of predicting the development of genetic contribution is that risks in breeding schemes, measured by parameters such as ΔF (Meuwissen and Woolliams 1994), cannot be described without a knowledge of the dynamics of individual contributions. The importance of the expected genetic contribution is made greater by the result of Woolliams (1998), which indicated that ΔF may be predicted from the expectation alone. The framework presented here provides a step-by-step recipe for predicting this expected genetic contribution over multiple generations. In providing the results, particular approaches have been described to derive the necessary regression models (appendices a and b). In other situations, such as the use of quadratic indices (Meuwissen 1997; Grundy *et al.* 1998), the formulae given in the appendices, based upon truncation selection, may not be appropriate whereas the results given in Equations 7, 9, 10, and 11 may remain valid. Therefore it is important to recognize that the details of these appendices are not an integral part of the recipe and other approaches could replace them in the recipe to suit the needs of a particular study.

J.A.W. gratefully acknowledges the Ministry of Agriculture, Fisheries and Food (United Kingdom) for funding this work, the encouragement of Dr. P. M. Visscher and Dr. B. J. McGuirk, and Professor A. Maki-Tanila and Professor B. Kinghorn for providing opportunities to develop and complete it. P.B. gratefully acknowledges financial support from the Netherlands Technology Foundation coordinated by the Earth and Life Science Foundation, and B.V. gratefully acknowledges financial support from the Biotechnology and Biological Sciences Research Council.

LITERATURE CITED

Bichard, M., A. H. R. Pease, P. H. Swales and K. Özkütük, 1973 Selection in a population with overlapping generations. Anim. Prod. **17:** 215–227.

Bijma, P., and J. A. Woolliams, 1999 Prediction of genetic contributions and generation intervals in populations with overlapping generations under selection. Genetics **151:** 1197–1210.

Bijma, P., J. A. M. Van Arendonk and J. A. Woolliams, 1999 Prediction of rates of inbreeding in populations with overlapping generations under mass selection. Genetics (in press).

Bulmer, M. G., 1980 *The Mathematical Theory of Quantitative Genetics*. Clarendon Press, Oxford.

Grundy, B., B. Villanueva and J. A. Woolliams, 1998 Dynamic selection procedures for constrained inbreeding and their consequence for pedigree development. Genet. Res. **72:** 159–168.

Hill, W. G., 1974 Prediction and evaluation of response to selection with overlapping generations. Anim. Prod. **18:** 117–139.

James, J. W., 1977 A note on the selection differential and generation length when generations overlap. Anim. Prod. **24:** 109–112.

James, J. W., and G. McBride, 1958 The spread of genes by natural and artificial selection in a closed poultry flock. J. Genet. **56:** 55–62.

Meuwissen, T. H. E., 1997 Maximizing the response of selection with a predefined rate of inbreeding. J. Anim. Sci. **75:** 934–940.

Meuwissen, T. H. E., and J. A. Woolliams, 1994 Response versus risk in breeding schemes. Proc. 5th World Congr. Genet. Appl. Livestock Prod. **18:** 236–243.

Rendel, J. M., and A. Robertson, 1950 Estimation of genetic gain in milk yield by selection in a closed herd of dairy cattle. J. Genet. **50:** 1–8.

Verrier, E., J. J. Colleau and J. L. Foulley, 1993 Long term effects of selection based on the animal model BLUP in a finite population. Theor. Appl. Genet. **87:** 446–454.

Villanueva, B., and J. A. Woolliams, 1997 Optimization of breeding programmes under index selection and constrained inbreeding. Genet. Res. **69:** 145–158.

Villanueva, B., J. A. Woolliams and B. Gjerde, 1996 Optimum designs for breeding programmes under mass selection with an application in fish breeding. Anim. Sci. **63:** 563–576.

Woolliams, J. A., 1998 A recipe for the design of breeding schemes. Proc. 6th World Congr. Genet. Appl. Livestock Prod. **25:** 427–430.

Woolliams, J. A., and R. Thompson, 1994 A theory of genetic contributions. Proc. 5th World Congr. Genet. Appl. Livestock Prod. **19:** 127–134.

Woolliams, J. A., N. R. Wray and R. Thompson, 1993 Prediction of long-term contributions and inbreeding in populations undergoing mass selection. Genet. Res. **62:** 231–242.

Woolliams, J. A., P. Bijma and B. Villanueva, 1999 Applications

and additional appendices to "Expected genetic contributions and their impact on gene flow and genetic gain." http://www.ri.bbsrc.ac.uk/geneflow/ (1 May 1999).
Wray, N. R., and R. Thompson, 1990 Prediction of rates of inbreeding in selected populations. Genet. Res. **55:** 41–54.
Wray, N. R., J. A. Woolliams and R. Thompson, 1990 Methods for predicting rates of inbreeding in selection populations. Theor. Appl. Genet. **80:** 503–512.
Wray, N. R., J. A. Woolliams and R. Thompson, 1994 Prediction rates of inbreeding in populations undergoing index selection. Theor. Appl. Genet. **87:** 878–892.

Communicating editor: R. G. Shaw

APPENDIX A: A GENERAL APPROXIMATION TO λ_{pq}

The regression of selection score of the unselected candidates of category p on the index I is given by $\omega_p \iota_p / \sigma_I$ (Wray and Thompson 1990), where ω_p is the selection proportion for category p. For a parent i of category q, the regression of the candidate index on $\boldsymbol{s}_{i(q)}$ for all the parents of category p that are of the same sex as category q was derived by standard index theory appropriate to the inheritance model under consideration (denote the coefficients for the regression on $(\boldsymbol{s}_{i(q)} - \bar{\boldsymbol{s}})$ by $\boldsymbol{w}$).

For each offspring of the parent from group q the probability of selection can then be approximated by $\omega_p(1 + \iota\sigma_I^{-1}\boldsymbol{w}^T(\boldsymbol{s}_{i(q)} - \bar{\boldsymbol{s}}))$. The expected number of offspring for a parent of category p is then $n_p\omega_p(1 + \iota\sigma_I^{-1}\boldsymbol{w}^T(\boldsymbol{s}_{i(q)} - \bar{\boldsymbol{s}}))$, where n_p is the number of candidates in category p per parent. $n_p\omega_p$ is equal to or $2g_{0,pq}X_pX_q^{-1}$, where g_0 is the proportion of genes among the newborn category p that derive from category q.

Considering only category q parents, they have an average selective advantage given by $\bar{\boldsymbol{s}}_q$ so the expectation is $2g_{0,pq}X_pX_q^{-1}(1 + \iota\sigma_I^{-1}\boldsymbol{w}^T(\boldsymbol{s}_{i(q)} - \bar{\boldsymbol{s}}_q) + \iota\sigma_I^{-1}\boldsymbol{w}^T(\bar{\boldsymbol{s}}_q - \bar{\boldsymbol{s}}))$. For sufficiently small deviations this is $\sim 2g_{0,pq}X_pX_q^{-1}(1 + \iota\sigma_I^{-1}\boldsymbol{w}^T(\boldsymbol{s}_{i(q)} - \bar{\boldsymbol{s}}_q))(1 + \iota\sigma_I^{-1}\boldsymbol{w}^T(\bar{\boldsymbol{s}}_q - \bar{\boldsymbol{s}}))$, where the last term in the product may be viewed as the additional selective advantage of category q, and so $g_{0,pq}(1 + \iota\sigma_I^{-1}\boldsymbol{w}^T(\bar{\boldsymbol{s}}_q - \bar{\boldsymbol{s}})) \approx g_{pq}$ and $\lambda_{pq} \approx \iota\sigma_I^{-1}\boldsymbol{w}$.

APPENDIX B: DERIVATION OF π_{pq}

Let $\boldsymbol{s}_{i(q)}$ be the vector of deviations of explanatory variables from their mean for a parent in category q and $\boldsymbol{s}_{j(p)}$ for an unselected progeny in category p and likewise $I_{j(p)}$ be the index upon which will be decided the selection, or otherwise, of $j(p)$. Let $\boldsymbol{s} = (\mathrm{s}_{i(q)}^T|\mathrm{s}_{j(p)}^T|I_{j(p)})$ have the partitioned (co)variance matrix

$$\begin{pmatrix} \boldsymbol{V}_{qq} & \boldsymbol{V}_{pq}^{\mathrm{T}} & \boldsymbol{\nu}_q \\ \boldsymbol{V}_{pq} & \boldsymbol{V}_{pp} & \boldsymbol{\nu}_p \\ \boldsymbol{\nu}_q^{\mathrm{T}} & \boldsymbol{\nu}_p^T & \sigma_I^2 \end{pmatrix}.$$

Before selection among candidates in category p, $\boldsymbol{s}_{i(q)}$ and $\boldsymbol{s}_{j(p)}$ can be expressed as regressions on $\boldsymbol{I}_{j(p)}$:

$$\boldsymbol{s}_{j(p)} = \sigma_I^{-2}\boldsymbol{\nu}_p I_{j(p)} + \boldsymbol{\varepsilon}_{j(p)}$$

$$\boldsymbol{s}_{i(q)} = \sigma_I^{-2}\boldsymbol{\nu}_q I_{j(p)} + \boldsymbol{\varepsilon}_{i(q)}.$$

Equating $E[\boldsymbol{s}_{i(q)}\boldsymbol{s}_{i(q)}^T]$ to $\boldsymbol{V}_{qq}$ gives $E[\boldsymbol{\varepsilon}_{i(q)}\boldsymbol{\varepsilon}_{i(q)}^T] = \boldsymbol{V}_{qq} - \sigma_I^{-2}\boldsymbol{\nu}_q\boldsymbol{\nu}_q^T$ and, similarly, $E[\boldsymbol{\varepsilon}_{j(p)}\boldsymbol{\varepsilon}_{i(q)}^T] = \boldsymbol{V}_{pq} - \sigma_I^{-2}\boldsymbol{\nu}_p\boldsymbol{\nu}_q^T$. After selection, Normal distribution theory infers that the regression coefficients on $I_{j(p)}$ are unchanged, but other regression coefficients are changed. Therefore, after selection

$$\boldsymbol{V}_{pq}^* = (\boldsymbol{V}_{pq} - k_p\sigma_I^{-2}\boldsymbol{\nu}_p\boldsymbol{\nu}_q^T), \quad \boldsymbol{V}_{qq}^* = (\boldsymbol{V}_{qq} - k_p\sigma_I^{-2}\boldsymbol{\nu}_q\boldsymbol{\nu}_q^T).$$

Let $\boldsymbol{\pi}_{pq}$ be the matrix of coefficients of $\boldsymbol{s}_{j(p)}$ on $\boldsymbol{s}_{i(q)}$ after selection; then $\boldsymbol{\pi}_{pq} = \boldsymbol{V}_{pq}^*\boldsymbol{V}_{qq}^{*-1}$.

In the applications described this is approximated by $\boldsymbol{\pi}_{pq} = \boldsymbol{V}_{pq}^*\boldsymbol{V}_{qq}^{-1}$. This is for three reasons: (i) simpler forms; (ii) it coincides with preceding published theory on genetic contributions; and (iii) such an assumption is implicit in standard index theory.

As an example with more than a single variable consider mass selection in discrete generations with random mating, where the vector of selective advantages explicitly includes the breeding value of the mate as well as the individual. There are two categories, males and females. In this case $\boldsymbol{s}_{i(q)}$ has two variables for each parent in category q, $(A_{i(q)} - \bar{A}_q, A_{i(q')} - \bar{A}_{q'})$, where $A_{i(q)}$ is the breeding value of i in category q, and $A_{i(q')}$ is the breeding value of its mate, and define $\boldsymbol{s}_{j(p)}$ similarly for the selected progeny $j(p)$. $\boldsymbol{V}_{pq} = (\frac{1}{2}\sigma_A^2(1 - k_qh^2), \frac{1}{2}\sigma_A^2(1 - k_{q'}h^2)|0, 0)$, $\boldsymbol{\nu}_p = (\sigma_A^2|0)$, $\boldsymbol{\nu}_q = (\frac{1}{2}\sigma_A^2(1 - k_qh^2)|\frac{1}{2}\sigma_A^2(1 - k_{q'}h^2))$, $\boldsymbol{V}_{qq} = \mathrm{diag}(\sigma_A^2(1 - k_qh^2), \sigma_A^2(1 - k_{q'}h^2))$, resulting in $\boldsymbol{\pi}_{pq} = (\frac{1}{2}(1 - k_ph^2), \frac{1}{2}(1 - k_ph^2)|0, 0)$. These are results of Wray and Thompson (1990). This example was chosen to obtain a fuller description of the expected long-term contribution by explicitly including the mate. Ignoring the mate is valid for considering genetic gain, providing the matrices are appropriately constructed; *e.g.*, if mating had been assortative rather than random the covariance between parent and offspring breeding value would need to account for the mate implicitly.

APPENDIX C: CONTRIBUTIONS OVER FINITE TIME WHEN α IS ESTIMATED AS A RIGHT EIGENVECTOR OF $(\boldsymbol{G}^T + (\boldsymbol{G}^T \otimes \boldsymbol{D}^T)(\boldsymbol{I} - \boldsymbol{G}^T \otimes \Pi^T)^{-1}(\boldsymbol{G}^T \otimes \Lambda^T))$

Adjustment of Equations 7 is done assuming, for simplicity, the only selective advantage is the breeding value. For category q, a *selected* individual at time 0, the vector of contributions to *selected* individuals in categories at time t is given by $\boldsymbol{c}_q(t) + \boldsymbol{b}_q(t)\,(A_{i(q)} - \bar{A}_q)$.

The approach taken is to use a modified form of Equation 4:

$$r_{i(q)}(t) = \tfrac{1}{2} \sum_{\text{offspring } j \in \text{ category } p} r_{j(p)}(t - \text{age}(p)).$$

Therefore the expected contribution after t cohorts is calculated by considering the expected contributions of selected offspring in category p, for $t - \text{age}(p)$ cohorts.